PNAS RESEARCH ARTICLE | SOCIAL SCIENCES OPEN ACCESS

Socioeconomic inequalities in molecular risk for chronic diseases observed in young adulthood

Michael J. Shanahan[a,b,1], Steven W. Cole[c], Sudharshan Ravi[a,1], Justin Chumbley[a], Wenjia Xu[a], Cecilia Potente[a], Brandt Levitt[d], Julien Bodelet[a], Allison Aiello[d,2], Lauren Gaydosh[e], and Kathleen Mullan Harris[d,f,1]

Contributed by Kathleen Mullan Harris; received March 18, 2021; accepted August 31, 2022; reviewed by Michael McFarland, Lauren Schmitz, and Carol Worthman

Many common chronic diseases of aging are negatively associated with socioeconomic status (SES). This study examines whether inequalities can already be observed in the molecular underpinnings of such diseases in the 30s, before many of them become prevalent. Data come from the National Longitudinal Study of Adolescent to Adult Health (Add Health), a large, nationally representative sample of US subjects who were followed for over two decades beginning in adolescence. We now have transcriptomic data (mRNA-seq) from a random subset of 4,543 of these young adults. SES in the household-of-origin and in young adulthood were examined as covariates of *a priori*-defined mRNA-based disease signatures and of specific gene transcripts identified *de novo*. An SES composite from young adulthood predicted many disease signatures, as did income and subjective status. Analyses highlighted SES-based inequalities in immune, inflammatory, ribosomal, and metabolic pathways, several of which play central roles in senescence. Many genes are also involved in transcription, translation, and diverse signaling mechanisms. Average causal-mediated effect models suggest that body mass index plays a key role in accounting for these relationships. Overall, the results reveal inequalities in molecular risk factors for chronic diseases often decades before diagnoses and suggest future directions for social signal transduction models that trace how social circumstances regulate the human genome.

social inequality | social genomics | biodemography | life-span development | social epidemiology

Significance

The analysis of gene expression in peripheral whole blood of US young adults in their late 30s revealed socioeconomic status-based inequalities in the molecular underpinnings of the most common chronic conditions of aging. Associations involved immune, inflammatory, ribosomal, and metabolic pathways, and extra- and intracellular signaling. Body mass index was a plausible, sizable mediator of many associations. Results point to new ways of thinking about how social inequalities "get under the skin" and also call for renewed efforts to prevent chronic conditions of aging decades before diagnoses.

Contemporary cohorts of Americans in early adulthood, spanning the third and fourth decades of life, are generally disease-free but notably at risk for debilitating conditions in the years to follow (1). Indeed, the prevalence of many common chronic conditions that characterize older adults—including cardiovascular disease (CVD), rheumatoid arthritis, chronic obstructive pulmonary disease (COPD), and Alzheimer's disease—increase markedly from the 40s onward (2–7). By age 65, almost seven in ten Americans have two or more such diagnoses (8). This health pattern of aging, reflecting the development of precursor risk in young adulthood and later diagnosis, illustrates how decades of social and biological wear and tear can eventuate in chronic disease states (9, 10).

Such wear and tear, however, involving diverse forms of stressors, is not randomly distributed in the population. Indeed, common adult diseases are characterized by a socioeconomic status (SES) gradient, comprising health differences by education, income, occupation, and subjective social status (11–17). In studies of older adults (typically 50–70 y old), increasing childhood and adult SES are negatively associated with mechanisms that drive many diseases of adulthood and that are indicated by gene expression patterns in circulating peripheral blood. Pro-inflammatory action is most commonly observed among people with low SES backgrounds (18–23), with additional evidence for suppressed antiviral response (24). This pattern of gene regulation has been described as a "conserved transcriptional response to adversity" (CTRA), a characteristic cross-species reaction to sudden stressors (25) that, if activated for prolonged periods, negatively impacts health (26).

Among young adults, low SES in the household-of-origin has been linked to a "defensive phenotype" characterized by up-regulated transcription of genes controlled by the CREB/ATF and nuclear factor κB (NF-κB) transcription control pathways and down-regulated activity of genes controlled by the glucocorticoid receptor, indicating pro-inflammatory action (18) that is consistent with the CTRA model. Whole-genome studies also observe associations between adult SES and a wide range of other disease-generating mechanisms, including extracellular signaling and cellular differentiation (23, 27). However, little is known about socioeconomic inequalities in the molecular precursors of late life diseases in early adulthood, before those diseases become prevalent [see (18) for an initial, targeted analysis]. In the present study, we used a combination of disease-based hypothesis-testing and unbiased discovery to map socioeconomic variations

Author affiliations: [a]Jacobs Center for Productive Youth Development, University of Zürich, Zürich, CH 8050; [b]Department of Sociology, University of Zürich, Zürich, CH 8050; [c]School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095; [d]Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27516; [e]Department of Sociology, University of Texas at Austin, Austin, TX 78712; and [f]Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3210

Author contributions: M.J.S., S.W.C., and K.M.H. designed research; M.J.S., W.X., L.G., and K.M.H. performed research; S.R., J.C., W.X., C.P., B.L., and J.B. analyzed data; and M.J.S., S.W.C., S.R., J.C., C.P., A.A., L.G., and K.M.H. wrote the paper.

Reviewers: M.M., Florida State University; L.S., University of Wisconsin-Madison; and C.W., Emory University.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: michael.shanahan@jacobscenter.uzh.ch or sudharshan.ravi@jacobscenter.uzh.ch or kathie_harris@unc.edu.

[2]Present address: Robert N. Butler Columbia Aging Center and Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY 10032.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2103088119/-/DCSupplemental.

Published October 17, 2022.

in the molecular precursors to the common chronic diseases in a well-powered sample of young adults in the United States. These analyses provide an unprecedented opportunity to discover the molecular pathways through which socioeconomic disparities in health in later life are structured by socioeconomic influences on molecular function decades earlier.

We draw on multiple indicators of SES in both the household-of-origin and in young adulthood and mRNA-based signatures of these chronic conditions in a sample of 4,543 young adults participating in the National Longitudinal Study of Adolescent to Adult Health (Add Health) (28). The data are derived from the largest nationally-representative study of young adults with blood transcriptomic data. The signatures reflect mRNA abundance levels of genes that have been empirically identified as markers of common, chronic adult diseases in genome-wide and genome-wide expression studies. Thus, the signatures describe the functional genomic action of diverse biological precursors to chronic conditions in a sentinel tissue.

The analyses focus on three descriptive tasks: 1) examining associations between standard indicators of SES in the household-of-origin and in young adulthood and *a priori*-defined disease signatures; 2) identifying, *de novo*, genes that are differentially expressed (DE) as a function of SES, as well as their biological significance; and 3) considering whether associations observed in steps (1) and (2) may be explained by key risk factors in the young adult population of the United States (body mass index [BMI], current smoking status, alcohol consumption, general perceived stress, difficulty paying bills, and health insurance). Previous research was insufficiently powered to test for putative mechanisms linking SES to gene expression profiles. Knowledge of such mechanisms may sharpen our understanding of life course patterns of molecular risk.

Results

Young Adult SES and Later Adult Disease Signatures. We begin by testing the hypothesis that at least one gene in a signature is significantly associated with indicators of SES (Fig. 1*A*) and then identify gene sets within signatures that are associated with SES, preserving the distinction between up- and down-regulation (Fig. 1*B*). In addition to the disease signatures, we also examined a signature for inflammation (designated "1KI") (29). Neither the household-of-origin SES composite nor any of its indicators were associated with any disease signature, and so parental SES is not considered further. Showing the results for adult SES among respondents who have not been diagnosed with these diseases, Fig. 1*A* reveals that, first, as expected, the 1KI was associated with the adult SES composite (t = –8.73, $P = 5.42 \times 10^{-14}$) and with education ($t = -5.77$, $P = 6.16 \times 10^{-5}$), income ($t = -6.73$, $P = 2.16 \times 10^{-7}$), subjective status ($t = -4.68$, $P = 6.16 \times 10^{-7}$), and occupation ($t = -6.26$, $P = 6.20 \times 10^{-6}$). Second, the adult SES composite was associated with all disease signatures: Alzheimer's (t = –5.29, $P = 2.01 \times 10^{-5}$) asthma ($t = -5.85$, $P = 3.92 \times 10^{-6}$), chronic kidney disease (CKD, $t = -7.21$, $P = 1.07 \times 10^{-9}$), COPD ($t = -5.93$, $P = 9.14 \times 10^{-7}$), CVD ($t = -4.62$, $P = 3.35 \times 10^{-4}$), depression ($t = -8.73$, $P = 5.42 \times 10^{-14}$), diabetes type 2 ($t = -3.82$, $P = 6.16 \times 10^{-3}$), hypertension ($t = -5.70$, $P = 5.60 \times 10^{-6}$), and rheumatoid arthritis ($t = -5.51$, $P = 7.57 \times 10^{-6}$). Finally, the young adult SES composite, income, occupation, and subjective status were the predominant correlates of the signatures.

Given the well-established emphasis on inflammatory pathways in extant research, we distinguished between patterns that reflect key inflammation-related genes (indicated by an inflammation signature) and other biological mechanisms. The patterns in Fig. 1*A* were observed when genes in the 1KI signature were excluded from the disease signatures (reported in *SI Appendix*, Fig. S1). The results reported in Fig. 1*A* were also recalculated for the entire mRNA sample, including subjects who reported diagnosis of at least one of the ten diseases (i.e., thus, the estimates may be biased by colliders). These results are reported in *SI Appendix*, Fig. S2 and suggest no notable changes in patterns reported in Fig. 1*A*. To shed light on the robustness of these associations, we present e-values for total

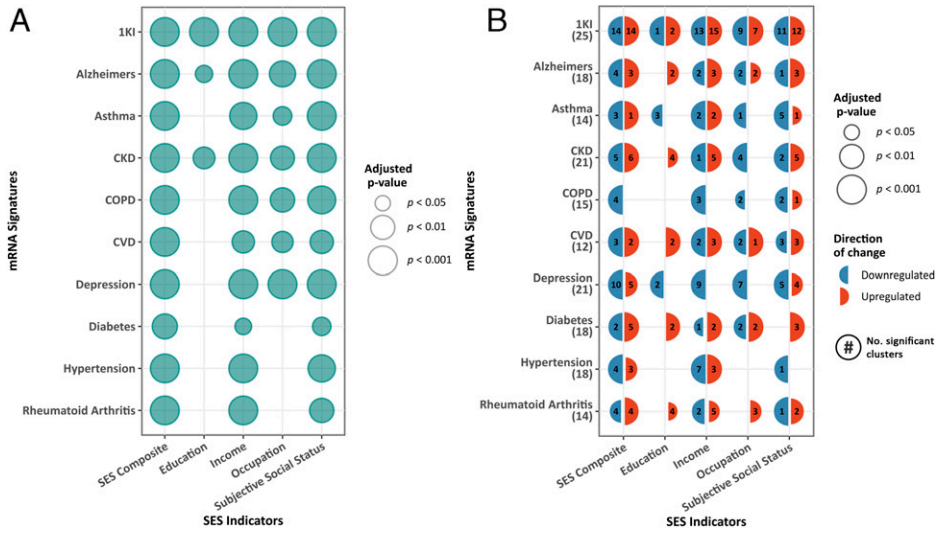

Fig. 1. Associations between SES and disease signatures. (*A*) Omnibus statistical significance (*P* on the −log10 scale) for associations between adult SES and disease signatures. The omnibus *P* values were calculated as the minimum (across all genes in the disease sets) FDR-corrected *P* value derived from a whole-genome (limma) linear regression for the association between SES and each gene in the signature with full controls. The reported significance thus indicates whether at least one gene in the disease set is predicted by SES, with type I errors corrected for the whole genome. (*B*) Number of up- and down-regulated clusters of genes associated with SES. The entire genome was clustered using WGCNA to identify clusters (25 in total). For each disease signature, its constituent genes were matched to the whole genome clusters and, within these clusters, to the up- or down-regulated subset. Within these groups, abundance levels were averaged. For example, 101 genes in the Alzheimer's signature belonged to 18 of the 25 whole genome clusters; of these 18, the SES composite was associated with the down-regulated portion of four clusters. Specific genes in these four clusters are reported in Dataset S1 (for further details, see *SI Appendix*, Fig. S12*B*).

effects in *SI Appendix*, Table S1. Results suggest that associations involving occupation are relatively susceptible to effects of unmeasured confounds.

Fig. 1*B* reports associations between indicators of SES and subsets of the disease signatures that overlap with the whole genome clusters (see *SI Appendix*, Fig. S12 *A* and *B* for analytic workflow). Enriched pathway analysis of the significantly predicted clusters revealed the functional significance of genes in significant clusters in Fig. 1*B*. The patterns in Fig. 1*B* were generally unchanged when genes in the 1KI signature were excluded from the disease signatures (*SI Appendix*, Fig. S1) The full Reactome results are reported in Dataset S1, and the figure and dataset suggest several conclusions. First, enriched pathways reflect a diverse range of biological processes that are both up- and down-regulated, suggesting that SES perturbs many biological systems involved in chronic disease in complex ways. The pathways associated with 1KI are illustrative. Most prominently, and consistent with past research, these pathways include manifold aspects of immunity and inflammation: for example, *HLA-* genes (human leukocyte antigen, associated with regulation of the immune system, infectious diseases, diabetes type II, and cancers); *CD-* genes (referring to immune-related cell surface signaling); *IRF-* genes (interferon regulatory factors involved in antiviral response); *IL-* genes (referring to interleukins); and many types of signaling (e.g., cytokine, *NOTCH-* and *MAPK-* genes). These same pathways may be down-regulated (albeit for different dimensions of SES), suggesting that associations between SES and mRNA activity are not uniformly up- and down-regulated for a given pathway.

Second, the SES composite, income, occupation, and subjective status are associated with specific disease signature clusters. Although different diseases often share common pathways, there is little, if any, overlap in specific genes in these shared pathways. Similarly, the same pathways may be enriched both for the 1KI and specific disease signatures, but there is no notable overlap in specific genes. Finally, the genes identified in Reactome pathways are typically highly multifaceted in their functions but a focus on patterns suggests that the functional significance of genes identified in Fig. 1*B* clusters reflect diverse aspects of immunity and inflammation, cell cycle (especially mitosis and apoptosis), transcription and translation, cell surface interactions, signaling, and metabolism (see Dataset S1).

SES and Whole-Genome Differential Expression. In addition to the disease and 1KI signatures, which were defined *a priori*, we also identified other empirical genome-wide transcriptomic correlates of young adult SES, income, education, occupation, and subjective status. DE genes were identified with a whole-genome (limma) linear regression for the association of SES with each gene with full controls with false discovery rate (FDR) correction of P values ($P < 0.05$) over genes not in the signatures; results are reported in *SI Appendix*, Fig. S3. The SES composite is associated with 121 DE genes (53 over-expressed, and 68 under-expressed), of which 42 are unique to the SES composite and 79 are shared with indicators of SES. Income and subjective status are associated with 57 and 56 genes, respectively, but virtually all of these genes are also identified by variation in the SES composite. SES in the household of origin was associated with few DE genes (19 genes) and so is not considered further. Volcano plots are shown in *SI Appendix*, Fig. S4.

To identify the specific functional pathways related to SES, we examined DE genes with Reactome; the results are summarized in Fig. 2 and details are reported in Dataset S2. Many pathways

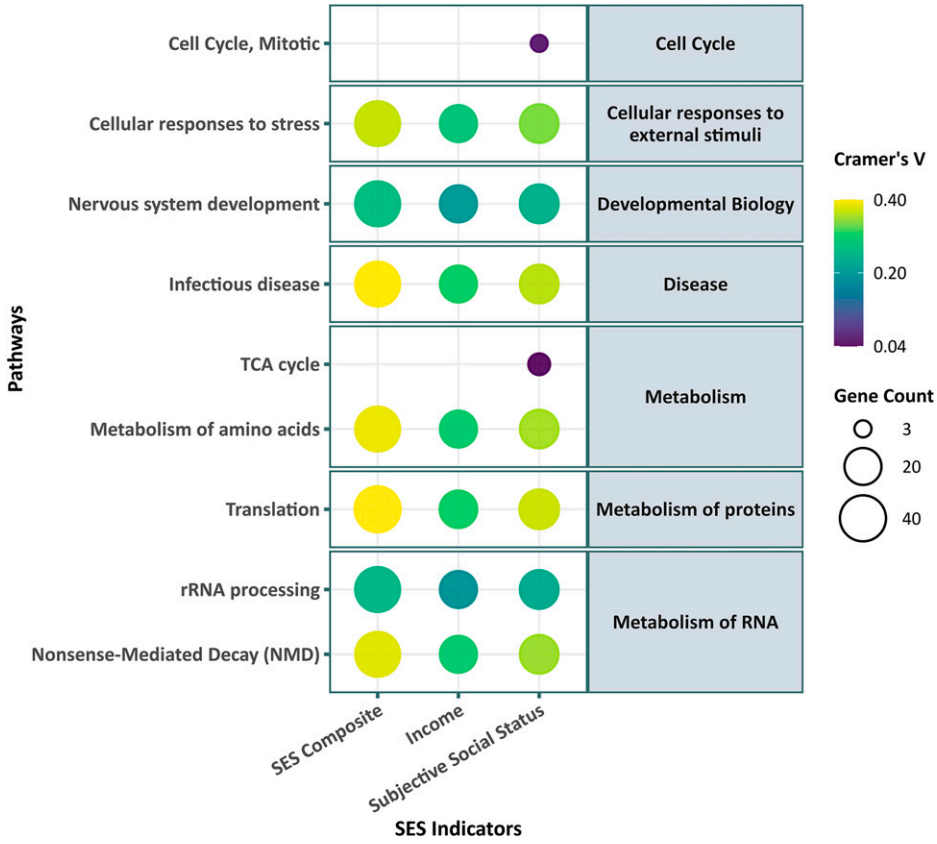

Fig. 2. Pathway enrichment of the *de novo* differentially expressed genes by adult SES. Significantly enriched Reactome pathways (with parent nodes reported to the *Right*, child nodes to the *Left*) for the differentially expressed genes by indicators of young adult SES, excluding genes from disease and 1KI signatures from Fig. 1. The size of the circle signifies the number of genes that contribute to the significant enrichment in a pathway and the color of the circle indicates Cramer's V, a measure of the magnitude of association.

shown in Fig. 2—including cellular response to stress, nervous system development, infectious disease, translation, nonsense-mediated decay, and rRNA processing—reflect a core of DE genes related to ribosomes and translation (a large number of *RPL-* and *RPS-* genes). These same genes figure prominently as part of a ribosomal subnetwork of a previously-identified senescence signature (30). The infectious disease pathway reflects genes induced by viral infections (e.g., influenza).

Genes associated with metabolism are prominent. Genes involved in the citric acid cycle (TCA) and respiratory electron transport chain are DE by the SES composite, income, and subjective social status. These pathways include genes encoding protein subunits essential for the proper functioning of mitochondrial oxidative phosphorylation. Dysregulation of the aerobic metabolism is accompanied by disruptions in the amino acid metabolism, predominantly by several genes involved in translation (*RPL-* and *RPS-* genes). Metabolic disruption, particularly in aerobic energy generation (TCA cycle and electron transport chain), is associated with a wide variety of diseases and may act as a primary source of age-related disorders (31, 32).

Risk Factors Associated with SES and Disease Signatures. Having examined SES correlates of mRNA signatures defined *a priori* as disease and 1K signatures and DE genes identified *de novo*, we asked whether commonly studied behavioral risk factors—BMI, current smoking status, alcohol consumption, general perceived stress, difficulty paying bills, and access to health insurance—could explain the observed associations. BMI most consistently accounted for a substantial part of the associations reported in Fig. 1*B* and also the *de novo* genes. Fig. 3 reports the proportion mediated by BMI for the significantly predicted clusters derived from the weighted correlation network analysis (WGCNA) of the whole genome (see Fig. 1*B* and *SI Appendix*, Fig. S12*B*). The proportion of mediation ranged from 8.1% (education and up-regulated diabetes) to 90.3% (SES composite and rheumatoid arthritis). In addition to the patterns involving BMI, we found that current smoker status emerged as a possible mediator between the SES composite and income and many disease signatures (*SI Appendix*, Fig. S5*B*). We then calculated the joint ratio of mediation (joint mediated effect/total effect) accounted for by all mediators simultaneously (using R package *multimediate*) for each cluster that significantly accounted for an SES-disease association (e.g., the percent of all mediators across three clusters that account for the SES composite-Alzheimer's). The results, reported in *SI Appendix*, Fig. S6, show that many of the percentages increase appreciably, although some results are no longer significant (reflecting a *P* value based on the joint mediation of all mediators).

The *de novo* genes were clustered in the same manner used with the signatures (Fig. 1*B*), resulting in 21 clusters that categorized the 121 DE genes. *SI Appendix*, Fig. S7 reports the mediational results for BMI, and *SI Appendix*, Fig. S8 reports the multiple mediator results. Patterns suggest that BMI could explain much of the associations between indicators of SES and SES-based clusters of *de novo* genes, but that the other mediators often increase the median proportion mediated considerably.

Finally, we examined e-values for the mediational models (see *SI Appendix*, Table S2) and results suggest that patterns involving the SES composite and occupation are relatively sensitive to unmeasured confounds.

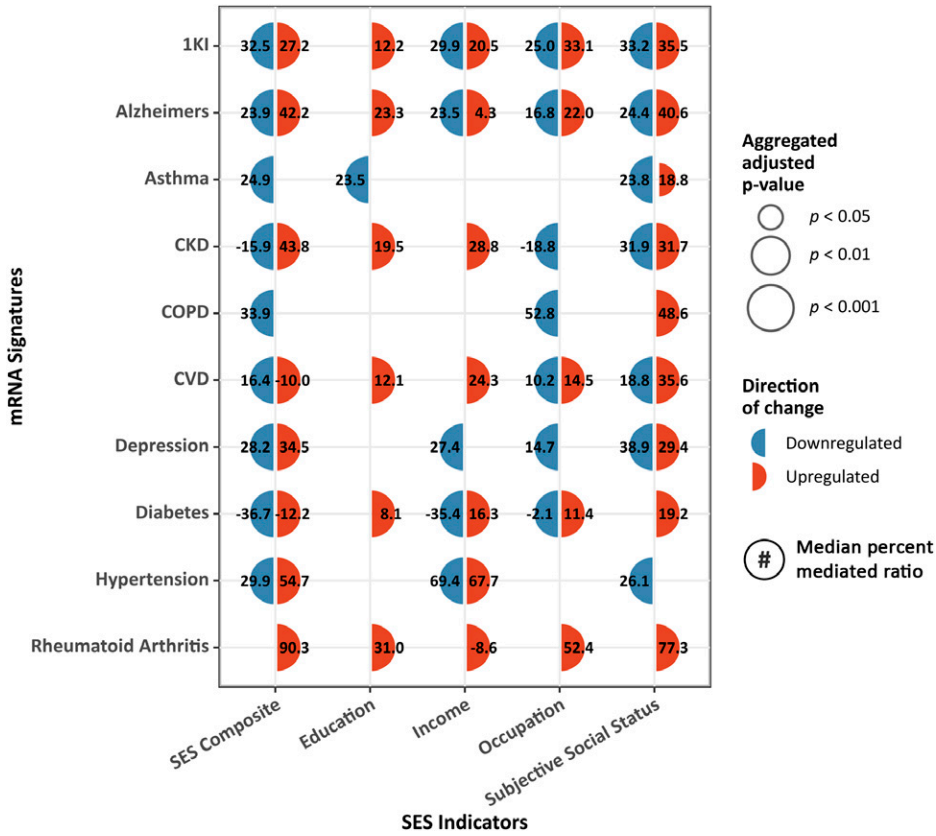

Fig. 3. Mediational models for BMI and significant clusters associated with disease signatures. Median percent mediated ratio (ACME/total effect) (superimposed number): average expression of significant clusters in Fig. 1*B* were used for mediational models. *P* values are corrected within column and significant mediational results for each signature combined using Fisher's method. The color scheme is the same as that in Fig. 1*B*. Negative values suggest a suppression pattern.

Discussion

The analyses revealed socioeconomic inequalities in young adults' molecular risk profiles for common chronic conditions of later adulthood in the United States and, additionally, in gene expression profiles associated with aging. Decades before these diseases are commonly observed in the population, their gene expression indicators show clear social status gradients, especially with respect to the young adult SES composite, income, and subjective status. These associations were generally observed regardless of the inclusion of genes in the 1KI inflammation signature, which is a proxy for key inflammation-related genes. These analyses also identified de novo SES gradients in gene sets extending beyond the traditionally emphasized functional pathways of inflammation and immunity. As such, the present results are consistent with prior research linking inflammation and immune regulation to social adversity (25, 33, 34), but they also identify additional biological domains through which SES might potentially impact health. Evidence suggested that SES-related differences in BMI contribute to many associations between SES and the signatures and genes identified *de novo*.

The chief commonalities with previous studies include the prediction of the *a priori*-defined 1KI inflammatory signature by SES and evidence showing the sensitivity of pathways involving *MAPK*-, *TLR*-, *IRF*-, and *IL*- genes (18, 35, 36), among others.

The present study also identified unique sets of status-graded genes and pathways. With respect to inflammation and immunity, human leukocyte antigen (*HLA*-) genes and *RIPK* genes, for example, were prominent. Additional biological pathways associated with SES included genes central to the metabolism of RNA (*RPL*- and *RPS*- genes), and genes involved in metabolism (e.g., *ALOX*-, *PTG*-, and *CSNK*-). Given that genes typically have diverse functions, these pathways are interpretive but we have emphasized patterns. Among the functions of many of these genes are diverse intra- and extracellular signaling mechanisms (23, 27), suggesting that SES may be involved in key aspects of social signal transduction.

Two functional pathways are especially noteworthy: the ribosomal protein-encoding genes and genes associated with metabolism, both because of their central role in senescence and diseases of aging. These pathways have not previously been noted as a correlate of SES, and further study of overlap between genes related to SES and senescence is warranted. Intriguingly, reduced expression of these ribosomal genes characterizes a transcriptomic aging signature (30). In the present case, we observed down-regulation of ribosomal protein genes (although not the same specific genes in the Peters senescence network) with increasing SES, a dysregulative pattern that calls for further study but that may stem from an association between low SES and increased ribosome metabolic activity associated with increased BMI. Relatedly, several of the associations are inverse to what might be expected. The abundance levels of composite scores and specific genes are best interpreted as indicators of disease-related gene activity. That is, we find SES-graded differential expression in genes that have been specified *a priori* but in some instances the direction of association may be not as expected or the expected direction may not be known (e.g., if the genes were identified by genetic polymorphism associations with disease in the absence of transcript abundance measures). Future research will be required to identify the specific genes involved in specific disease etiologies at different points in life.

The results also depart from previous research insofar as they show little evidence that SES in the household-of-origin predicted disease expression profiles. Previous studies have reported relationships between household SES and inflammation/immunity, although in different national contexts (Canada, Italy, and Switzerland) and with different measures of parental SES (18, 19). We recoded our household occupational data to examine manual versus nonmanual paternal occupation (19), but the null findings with respect to occupation and disease signatures were robust.

Most prior research on SES and molecular risk draws on social signal transduction models based on stress biology (26, 37, 38), which suggests that social gradients in health stem at least in part from differential exposures and vulnerabilities to stressors (predominantly chronic psychosocial stressors). Although the present findings are consistent with this hypothesis in some respects (see *SI Appendix*, Fig. S5*C*), our research suggests additional transducing paths. Some of the associations between SES and disease-related gene expression may not involve stressors per se but rather income, education, and prestige as "flexible resources" that influence health in myriad ways beyond stress (39). Educational and financial resources are often needed to maintain one's health via, for example, health literacy, access to fresh foods and preventive medical care, housing in areas with relatively low levels of toxicants, ready access to green spaces for exercise, and inclusion in social networks of people with resources (40).

Indeed, the observed SES gradients were most consistently accounted for by BMI, which likely reflects the widespread damaging effects of subcutaneous adipose tissue on physiology, including inflammation (41, 42). Although stress likely contributes to adiposity (e.g., via "stress eating"), other SES-graded social and environmental mechanisms (e.g., food ecologies, etc.) also play substantial roles, possibly including a conserved response to status involving appetite (33). An expansion of current social signal transduction models beyond stressors to include other mechanistic pathways, especially obesogenic ones, is warranted.

Our study is unique for its large, diverse, and representative sample that includes standard measures of SES in the household-of-origin and in young adulthood and mRNA abundance levels from circulating whole blood. However, several limitations should be noted. First, experimental studies of rhesus macaques, a species that lives in social hierarchies and is closely related to humans, have shown that the distinct cell populations that constitute peripheral blood mononuclear cells condition gene expression (34). The sensitivity of specific cell types (e.g., natural killer cells) to status needs to be mapped out with greater precision in humans. Second, while the data allow for multivariate descriptive conclusions, the basis for causal inference is circumscribed, and the results should be construed in descriptive terms (see *SI Appendix*, Fig. S11 for major assumptions needed for causal inference). Although strategies exist to identify the effects of income, education, and occupation individually, such approaches are often limited in terms of their external validity. Third, although our measures of SES are standard operationalizations, they do not fully account for the complexity of SES, which likely includes compositional effects (that directly reflect inequality) and intersectional patterns involving sex, race, and ethnicity. The study of such nuances will require larger samples. Additionally, our study design does not allow us to disentangle the importance of young adulthood from that of recency, which is perfectly confounded in this sample. The repeated collection of mRNA in a large, diverse panel study—beginning earlier in the life course—would facilitate a better understanding of the origins of these forms of molecular risk. Finally, individual mRNA molecules are highly

transient, which may create stochastic patterns with considerable noise, and effect sizes are small; on the other hand, biological models of chronic stress recognize sources of stability in average mRNA abundance that derive from relatively stable patterns of gene regulation and *de novo* transcription (26). The present results reflect an additional stabilizing factor in that they are based on groups of genes, specified *a priori*, that have well-established significance for human health (43).

The present study reports SES-based inequalities in molecular risk factors for common chronic diseases of later adulthood in the decades before such conditions are diagnosed with frequency, and the analyses suggest mechanistic hypotheses for SES-related inequalities. Disparities in diseases of late adulthood likely reflect disparities in molecular risk in young adulthood. This risk involves not only inflammatory and immune pathways but also other biological pathways, and our *de novo* results highlight metabolic, ribosomal, and diverse signaling pathways as important targets for future mechanistic research on health disparities. Moreover, among the common risk mechanisms studied here, BMI consistently emerged as a plausible mediator of these associations. The results highlight the need for multifaceted policies and interventions that target stressors and obesogenic mechanisms early in life, before these socially based mechanisms eventuate in clinically relevant, costly health impairments.

Materials and Methods

Data. The data for this study are from Add Health, a representative study of US adolescents in grades 7–12 in 1994–1995 (age range, 12–18; mean age, 15.3; SD = 1.6) who were followed into adulthood over five waves of data collection (28). Participants in Wave V of Add Health consented to provide an intravenous blood sample in PAXgene RNA tubes. These samples were collected in 2016–2017 (44). The data for the present study include 4,543 transcriptomic profiles (based on mRNA-seq data; age range, 33–43; mean age, 37.33; SD = 1.85). Assuming an average 1.1-fold change among 20% DE genes (80% assumed not DE), an analysis of power (using R package *ssizeRNA*) indicated that a sample of 971 would be sufficient to attain average power of 0.8 while controlling FDR to 0.05 over the genome. *SI Appendix*, Fig. S9 shows how the analytic sample was derived, and *SI Appendix*, Table S3 documents representativeness relative to the Wave V total sample (n = 7,769). The mRNA sample is somewhat more female and higher on SES (reflecting education and occupation).

Measures. Transcriptome profiles were derived by 3′ end sequencing whole blood polyadenylated RNA (for a full description, see *SI Appendix, Study Protocol*). The resulting mRNA abundance levels from circulating whole blood were then used to compute composite scores based on previously defined disease signatures. *SI Appendix*, Table S4 presents the sources and properties of the empirically-derived disease signatures, and the list of signatures is shown in Fig. 1*A*. Given the central role of inflammation in many chronic diseases and its prominence in past social genomic research, we also included the 1K inflammation signature (1KI) (29). Analyses were conducted with and without the 1KI gene set, and differences between analyses are a heuristic approximation of the relative importance of central, inflammation-related genes. There was very little overlap between the ten disease signatures in terms of their constituent genes, but there was some overlap with the 1KI signature. Average correlations among disease signatures ranged from r = 0.14–0.29 but were much smaller in magnitude with the exclusion of 1KI genes from them.

Socioeconomic status composites, for both young adults and their parents, represent the sum of the standardized indicators. Education represents the highest self-reported completed years of education (averaged in cases of two parents). Income is the gross household income, log-transformed (for parents) and reported on an ordinal scale (for young adults) with values representing midpoints of categories and the highest value equaling the lowest value for that category ($200,000). Occupation represents the socioeconomic index score of parents' jobs (highest of two parents) and young adult's current job (45, 46). Subjective status in young adulthood was assessed using the MacArthur Scale of Subjective Social Status, which asks respondents to view a ten-rung ladder as representative of education, prestige, and money and to pick the step that shows where they think they currently stand relative to other people in the United States (47). The Wave V young adult SES composites with and without subjective status were highly correlated (r = 0.96). Correlations between indicators and composites of SES are shown in *SI Appendix*, Fig. S10.

Models. We examined quantitative variations in the abundance levels of genes in these signatures as a linear function of 1) socioeconomic composites from the household-of-origin or young adulthood and their constituent indicators (education, income, occupational education, and subjective status [assessed in young adulthood only]) and 2) covariates that are likely to influence mRNA abundance levels, including sex, race, age, pregnancy status, plate, number of hours fasting before blood sampling, use of anti-inflammatory medications in the past 4 wk (e.g., NSAIDS, Cox-2 inhibitors, inhaled corticosteroids), count of common subclinical symptoms in the past 4 wk (e.g., colds, flu, fever), and count of common infectious or inflammatory diseases in the past 4 wk (e.g., active infection, seasonal allergy), with correction for batch using ComBat. A directed acyclic graph depicting the generalized linear model is presented in *SI Appendix*, Fig. S11. The final models were assessed with and without the inclusion of people with self-reported diagnosed diseases, with the latter results reported in the main text (because this specification avoids collider bias) and the former reported in *SI Appendix*, Fig. S1.

We examined associations between indicators of SES and subsets of disease signatures that preserve the distinction between up- and down-regulated genes. The entire genome was clustered using WGCNA to identify clusters of co-expressed networks of genes (25 in total) (48). For each disease signature, we linked each of its genes to whole-genome clusters and, within each cluster, to its up- or down-regulated subset (i.e., 50 in total). We then averaged the abundance levels of genes for each subset. For further detail, see *SI Appendix*, Fig. S12*B*. Such a classification enabled us to find significant direction-specific gene sets for each disease signature that were associated with indicators of SES. This analytic strategy identifies the functional clusters across the genome in the sample and the relationships among these whole-genome clusters and genes in the disease signatures. An alternative strategy, which defines clusters within disease signatures, yielded similar results but is less readily interpretable because the number of clusters per signature varied considerably. Reliable coexpression networks cannot be identified within some disease signatures given their small size and the lack of other gene transcripts beyond the signature.

Finally, we performed a differential expression analysis for those genes not contained in the signatures. The functional significance of subsets of signatures and the DE genes was examined with pathway enrichment using Reactome. We also examined possible mediators of relationships that were observed between Wave V SES and disease signatures using a counterfactual mediational framework (49). The candidate mediators included commonly studied health risks: BMI, computed from measured height and weight; current smoking status (self-reported); alcoholic drinks on days drank over past 30 d [0 drinks, 1–2 drinks, and 3–5 drinks per occasion, more than 5, which is similar to a coding scheme that has been used in other papers (50)], perceived stress, based on the short version of Cohen's Perceived Stress Scale (51); and financial stress, assessed by asking respondents whether they had difficulty paying bills (see *SI Appendix*, Fig. S7 for correlation matrix) and access to health insurance (1 in case the respondent did not have health insurance because not offered, considered too expensive, or they did not want it, 0 otherwise). The average causal mediated effect (ACME) was estimated for each possible mediator. Viewed in causal terms, however, these models are premised on strong assumptions (see *SI Appendix*, Fig. S11), and thus the results are construed as multivariate descriptions that are hypothesis-generating.

A summary of the analytic pipeline is shown in *SI Appendix*, Fig. S12*A*. All analyses were conducted using R software, especially the Bioconductor suite (52), unless noted.

Data, Materials, and Software Availability. Add Health transcriptomic data are available via a restricted data contract. Additional information and application for the restricted-use data can be accessed through the Carolina

Population Center (CPC) data portal at https://data.cpc.unc.edu/projects/2/view (53). The rest of the Add Health data are not restricted and is available at https://www.cpc.unc.edu/projects/addhealth/documentation/ (54). The R code used for these analyses is available at https://github.com/socialgnome/PNAS-SES (55).

ACKNOWLEDGMENTS. This research was supported by NIH (R01-HD087061 to MPIs M.J.S. and K.M.H. [specifically for the present analyses], P30-AG017265, R01-AG043404, and R01-AG033590), by the Swiss National Science Foundation (10531C-197964 to M.J.S.), and by the Jacobs Center for Productive Youth Development (University of Zurich). This research uses data from Add Health, a program directed by Robert Hummer and designed by J. Richard Udry, Peter S. Bearman, and K.M.H. (University of North Carolina at Chapel Hill). The Add Health program is funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (P01-HD31921), with cooperative funding from 23 other federal agencies and foundations (https://addhealth.cpc.unc.edu/about/funders).

1. K. M. Harris, An integrative approach to health. *Demography* **47**, 1–22 (2010).
2. B. Doney *et al.*, Prevalence of airflow obstruction in U.S. adults aged 40-79 years: NHANES data 1988-1994 and 2007-2010. *COPD* **12**, 355–365 (2015).
3. G. A. Roth *et al.*; Global Burden of Cardiovascular Diseases Collaboration, The burden of cardiovascular diseases among US states: 1990-2016. *JAMA Cardiol.* **3**, 375–389 (2018).
4. A. H. Mokdad *et al.*, Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. *JAMA* **289**, 76–79 (2003).
5. C. G. Helmick *et al.*; National Arthritis Data Workgroup, Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. *Arthritis Rheum.* **58**, 15–25 (2008).
6. B. N. Vardarajan *et al.*; NIA-LOAD/NCRAD Family Study Group, Age-specific incidence rates for dementia and Alzheimer disease in NIA-LOAD/NCRAD and EFIGA families: National Institute on Aging Genetics Initiative for Late-Onset Alzheimer Disease/National Cell Repository for Alzheimer Disease (NIA-LOAD/NCRAD) and Estudio Familiar de Influencia Genetica en Alzheimer (EFIGA). *JAMA Neurol.* **71**, 315–323 (2014).
7. K. E. Barbour, C. G. Helmick, M. Boring, T. J. Brady, Vital signs: Prevalence of doctor-diagnosed arthritis and arthritis-attributable activity limitation–United States, 2013-2015. *MMWR Morb. Mortal. Wkly. Rep.* **66**, 246–253 (2017).
8. Centers for Medicare and Medicaid Services, *Chronic Conditions: Prevalence State/County Table: All Fee-for-Services Beneficiaries* (Centers for Medicare and Medicaid Services).
9. M. Chen, H. Zhu, Y. Du, G. Yang, How does the social environment during life course embody in and influence the development of cancer? *Int. J. Public Health* **63**, 811–821 (2018).
10. J. Lynch, G. D. Smith, A life course approach to chronic disease epidemiology. *Annu. Rev. Public Health* **26**, 1–35 (2005).
11. H. Choi *et al.*, Comparison of health outcomes among high- and low-income adults aged 55 to 64 years in the US vs England. *JAMA Intern. Med.* **180**, 1185–1193 (2020).
12. L. J. Samuel *et al.*, Socioeconomic disparities in six-year incident dementia in a nationally representative cohort of U.S. older adults: An examination of financial resources. *BMC Geriatr.* **20**, 156 (2020).
13. M. S. Pantell, A. A. Prather, J. M. Downing, N. P. Gordon, N. E. Adler, Association of social and behavioral risk factors with earlier onset of adult hypertension and diabetes. *JAMA Netw. Open* **2**, e193933 (2019).
14. A. S. Karlamangla, S. S. Merkin, E. M. Crimmins, T. E. Seeman, Socioeconomic and ethnic disparities in cardiovascular risk in the United States, 2001-2006. *Ann. Epidemiol.* **20**, 617–628 (2010).
15. K. L. Tang, R. Rashid, J. Godley, W. A. Ghali, Association between subjective social status and cardiovascular disease and cardiovascular risk factors: A systematic review and meta-analysis. *BMJ Open* **6**, e010137 (2016).
16. D. Gill, A. Efstathiadou, K. Cawood, I. Tzoulaki, A. Dehghan, Education protects against coronary heart disease and stroke independently of cognitive function: Evidence from Mendelian randomization. *Int. J. Epidemiol.* **48**, 1468–1477 (2019).
17. A. Karp *et al.*, Relation of education and occupation-based socioeconomic status to incident Alzheimer's disease. *Am. J. Epidemiol.* **159**, 175–183 (2004).
18. G. E. Miller *et al.*, Low early-life social class leaves a biological residue manifested by decreased glucocorticoid and increased proinflammatory signaling. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14716–14721 (2009).
19. R. Castagné *et al.*, Biological marks of early-life socioeconomic experience is detected in the adult inflammatory transcriptome. *Sci. Rep.* **6**, 38705 (2016).
20. T. Zeller *et al.*, Genetics and beyond–the transcriptome of human monocytes and disease susceptibility. *PLoS One* **5**, e10693 (2010).
21. M. E. Levine, S. W. Cole, D. R. Weir, E. M. Crimmins, Childhood and later life stressors and increased inflammatory gene expression at older ages. *Soc. Sci. Med.* **130**, 16–22 (2015).
22. S. Stringhini *et al.*, Life-course socioeconomic status and DNA methylation of genes regulating inflammation. *Int. J. Epidemiol.* **44**, 1320–1330 (2015).
23. N. Borghol *et al.*, Associations with early-life socio-economic position in adult DNA methylation. *Int. J. Epidemiol.* **41**, 62–74 (2012).
24. M. E. Levine, E. M. Crimmins, D. R. Weir, S. W. Cole, Contemporaneous social environment and the architecture of late-life gene expression profiles. *Am. J. Epidemiol.* **186**, 503–509 (2017).
25. S. W. Cole, The conserved transcriptional response to adversity. *Curr. Opin. Behav. Sci.* **28**, 31–37 (2019).
26. R. Nusslock, G. E. Miller, Early-life adversity and physical and emotional health across the lifespan: A neuroimmune network hypothesis. *Biol. Psychiatry* **80**, 23–32 (2016).
27. T. W. McDade *et al.*, Genome-wide analysis of DNA methylation in relation to socioeconomic status during development and early adulthood. *Am. J. Phys. Anthropol.* **169**, 3–11 (2019).
28. K. M. Harris *et al.*, Cohort profile: The National Longitudinal Study of Adolescent to Adult Health (Add Health). *Int. J. Epidemiol.* **48**, 1415–1415k (2019).
29. M. J. Loza *et al.*, Assembly of inflammation-related genes for pathway-focused genetic analysis. *PLoS One* **2**, e1035 (2007).
30. M. J. Peters *et al.*; NABEC/UKBEC Consortium, The transcriptional landscape of age in human peripheral blood. *Nat. Commun.* **6**, 8570 (2015).
31. R. Sharma, A. Ramanathan, The aging metabolome-biomarkers to hub metabolites. *Proteomics* **20**, e1800407 (2020).
32. S. van der Rijt, M. Molenaars, R. L. McIntyre, G. E. Janssens, R. H. Houtkooper, Integrating the hallmarks of aging throughout the tree of life: A focus on mitochondrial dysfunction. *Front. Cell Dev. Biol.* **8**, 594416 (2020).
33. B. K. Cheon, Y. Y. Hong, Mere experience of low subjective socioeconomic status stimulates appetite and food intake. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 72–77 (2017).
34. N. Snyder-Mackler *et al.*, Social status alters immune regulation and response to infection in macaques. *Science* **354**, 1041–1045 (2016).
35. S. W. Cole, Social regulation of human gene expression: Mechanisms and implications for public health. *Am. J. Public Health* **103** (suppl. 1), S84–S92 (2013).
36. E. Chen, G. E. Miller, M. S. Kobor, S. W. Cole, Maternal warmth buffers the effects of low early-life socioeconomic status on pro-inflammatory signaling in adulthood. *Mol. Psychiatry* **16**, 729–737 (2011).
37. S. W. Cole, Human social genomics. *PLoS Genet.* **10**, e1004601 (2014).
38. S. K. Lutgendorf, A. K. Sood, Biobehavioral factors and cancer progression: Physiological pathways and mechanisms. *Psychosom. Med.* **73**, 724–730 (2011).
39. J. C. Phelan, B. G. Link, P. Tehranifar, Social conditions as fundamental causes of health inequalities: Theory, evidence, and policy implications. *J. Health Soc. Behav.* **51** (suppl.), S28–S40 (2010).
40. B. G. Link, J. Phelan, Social conditions as fundamental causes of disease. *J. Health Soc. Behav.* **Spec No.**, 80–94 (1995).
41. M. Bradshaw, B. V. Kent, W. M. Henderson, A. C. Setar, Subjective social status, life course SES, and BMI in young adulthood. *Health Psychol.* **36**, 682–694 (2017).
42. P. H. Wirtz, U. Ehlert, L. Emini, T. Suter, Higher body mass index (BMI) is associated with reduced glucocorticoid inhibition of inflammatory cytokine production following acute psychosocial stress in men. *Psychoneuroendocrinology* **33**, 1102–1110 (2008).
43. S. W. Cole, Elevating the perspective on human stress genomics. *Psychoneuroendocrinology* **35**, 955–962 (2010).
44. S. W. Cole, M. J. Shanahan, L. Gaydosh, K. M. Harris, Population-based RNA profiling in Add Health finds social disparities in inflammatory and antiviral gene regulation to emerge by young adulthood. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 4601–4608 (2020).
45. R. M. Hauser, J. R. Warren, Socioeconomic indexes for occupations: A review, update, and critique. *Sociol. Methodol.* **27**, 177–298 (1997).
46. M. Hout, T. W. Smith, P. V. Marsden, "Prestige socioeconomic scores, 2010 Census codes" in *Methodological Report* (NORC, Chicago, 2016).
47. N. E. Adler, Health disparities: Taking on the challenge. *Perspect. Psychol. Sci.* **8**, 679–681 (2013).
48. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
49. D. Tingley, T. Yamamoto, K. Hirose, L. Keele, K. Imai, mediation: R package for causal mediation analysis. *J. Stat. Softw.* **59**, (2014).
50. R. Crosnoe, C. Riegle-Crumb, A life course model of education and alcohol use. *J. Health Soc. Behav.* **48**, 267–282 (2007).
51. S. Cohen, G. M. Williamson, "Perceived stress om a probability sample of the United States" in *The Social Psychology of Health*, S. Spacapan, S. Oskamp, Eds. (Sage, Newbury Park, CA, 1988), pp. 3–67.
52. R. Gentleman, *R Programming for Bioinformatics* (Chapman & Hall/CRC, London, 2008).
53. Carolina Population Center. Add heath. https://data.cpc.unc.edu/projects/2/view. Carolina Population Center. Accessed 28 September 2022.
54. Add Health. Documentation. https://addhealth.cpc.unc.edu/documentation/. Carolina Population Center. Accessed 28th September 2022.
55. M. J. Shanahan *et al.*, Socioeconomic inequalities in molecular risk for chronic diseases observed in young adulthood. GitHub. https://github.com/socialgnome/PNAS-SES. Deposited 10 March 2022.